

対話における文脈情報を用いた指示詞・形式名詞の解析

—バック旅行選択を例にして—

横田 聡一郎 武田紀子 飯田 善久

[yokota@sdsun9., takeda@iida@] is. seikei. ac. jp

成蹊大学工学部

1.はじめに

指示詞や形式名詞の照応に関する研究は、書き言葉を中心的対象として、長い間研究されてきており、表層的言語情報を用いる方法でも有効な処理ができることは報告されてきている。しかし、それだけでは意味的な処理がないため、理解や推論が必要である場合には解析が困難である。

一般的に、意味的な理解を含む解析を実現しようとすると、膨大な常識・知識が必要である。しかし、ある具体的な目的を持った小規模タスク対話においては、対話の内容は限られるので、知識の量を減らし、効率をあげることができる。本研究では、バック旅行選択対話を例にとり、意味的な文脈知識を利用して指示詞や形式名詞の照応を行うシステムを構築しているので、それについて報告する。

2.システムの概要

指示詞・形式名詞の指示対象を照応解析するシステムの全体の流れは、図1のようになる。

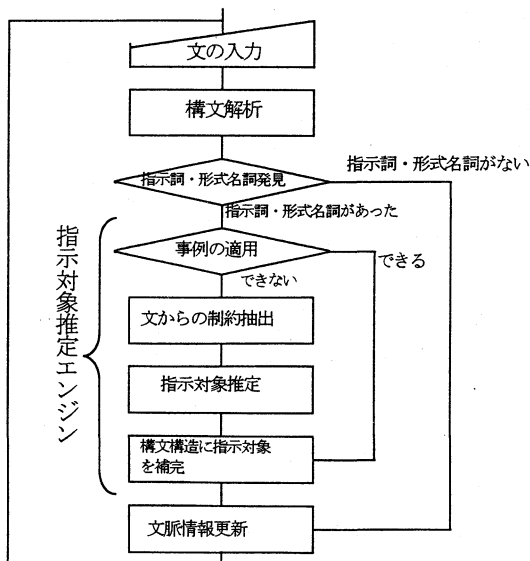


図1: システムの流れ

3.文脈テーブル

文脈処理を実現していくためには、参照すべき文脈モデルを構築する必要がある。図2のような構造をした対話の文脈テーブルを用意した。

この文脈テーブルが、文脈保持をするための基本的な役割を担い、1文（ひとつの述部を持つ文）ごとに、この文脈テーブルに情報が追加されていく。よって、述語接続助詞によって連結されたような複文は、構文解析の段階で2文に分解され、1文ずつ文脈テーブルに格納される。

| 1文ごとに格納 | |
|---------|----------------------------------|
| | 保持する情報 |
| (1) | 発話された文の肯定/否定 |
| (2) | 発話された文の種類 |
| (3) | 発話された文の述部にある動詞の意味 |
| (4) | 焦点となる語 |
| (5) | 焦点を含む名詞句 |
| (6) | 焦点となる語の意味 |
| (7) | 焦点とはならなかったが、次文で指示される可能性のある語とその意味 |
| (8) | ツアー名と焦点 |
| (9) | 焦点以外の語とツアー名 |
| (10) | 発話者 |
| (11) | 時間 |

図2: 文脈テーブルの構造

次に、文脈テーブルの構造の詳細を述べる。以下の説明文の番号と、テーブルの番号は対応している。

(2) 文の種類

バックツアー選択対話を含む、話し言葉による実際の旅行関連対話 105 例[1]を参考に、文の種類を以下の19種に分類した。

挨拶、依頼、希望、yes_no 質問、wh 系質問、確認、主張、同意、理由、提案、可能、推測、受領、返答、条件、断定、説明、過度、不明

この文の種類は、構文解析を行った後、述部として解析された部分の助動詞、終助詞、動詞の意味カ

カテゴリによって決定される。

(3) 述部の動詞の意味

動詞の意味カテゴリは、実際の対話例を参考にし、次の28種に分類した。

する、相談、行く、理解、見る、送る、必要、着く、決定、持つ、付く、かかる、泊まる、乗る、存在、申込、思う、検索、高い、安い、ものたりない、よい、ゆっくり、ちがひ、動作、移動、心理的、不明

(4) 焦点となる語

焦点は、発話された文において、話者が強調したいことを示すものである。焦点として特定された名詞、その名詞を含む名詞句、その名詞の意味カテゴリが文脈テーブルに格納される。焦点の特定方法については、次章で解説する。

(7) 焦点とならなかった語

ここには、文中に含まれる焦点化された名詞以外の名詞のうち、焦点化されることのできる意味カテゴリを持つ語と、その意味カテゴリが格納される。これは、次の文で、焦点よりは可能性が低い、指示される可能性のある語群である。文脈テーブルのうち、この部分までは、主に言語的文脈を保持する役割を持つ。

(8) ツアー名と焦点

焦点となる語に対応するツアー名とその項目が格納される。

(9) 焦点以外の語とツアー名

焦点とならない語に対応するツアー名とその項目が格納される。

4. 焦点特定の方法

焦点を特定することは、対話の言語的文脈を理解したり、指示詞・形式名詞の照応を行う上で非常に重要になる。本研究では、助詞に基づいた格情報と、文脈的なつながりに対応できる Centering アルゴリズム[2],[3]を基に、以下のような焦点特定の方法を考えた。

4.1 焦点特定アルゴリズム

焦点は、次のステップに従って決定される。

[1] 文中から、焦点候補のリストを取り出す。

- (1) 各名詞句内から、焦点となりうる意味カテゴリを持つ名詞をすべて取り出す。その際、名

詞句の主名詞を先に取り出す。

- (2) 取り出した名詞群を、意味カテゴリを基に優先度ランク1で並び替え、焦点候補のリストとして格納する。その時、同じ優先度のものが2つ以上あった場合は、優先度ランク2により並び替える。

優先度ランク1:

[ツアー、パンフレット、ツアー関連手続き、場所、予算、期間、イベント、日程、目的、相手、ホテル、食事、買物、気候、道具(含乗物)]

優先度ランク2:

1. 指示照応されているもの
2. 固有名詞
3. 助詞「は」でマークされているもの
4. 先に出現しているもの

- (3) 各名詞句内の主名詞以外の名詞を、(2)の優先度で格納していく。

- (4) 焦点候補リストの先頭を、主題とする。

[2] 焦点候補リストから、焦点を次の順番で決定する。

1. 述語動詞の必須格
2. 前文の焦点に対応するもの
3. 前文の焦点候補リストにあり、かつ、指示詞等で照応されたもの
4. 前文の焦点候補リスト中の対応する要素で、優先度の高いもの
5. 現在の文の後に述語接続助詞が続いている場合、述部の目的格に該当するもの
6. 指示詞等で照応されたもの
7. 固有名詞
8. 主題
9. 焦点候補リストが空であれば、焦点を-1とする。
この場合は、前文の焦点と焦点候補リストを保持して次文に渡す。

[3] 焦点候補リストから、焦点を除いて、「焦点とならなかった語群」とする。

5. ツアーに関する情報テーブル

対話の言語外文脈を保持するために、対話の目的となる「パックツアー」を中心としたツアー・テーブルと、申込者の希望を中心とした希望テーブルを用意した。それらのテーブルに含まれる基本情報は、図3のようなものである。

| | |
|--------|--------------------------------|
| パンフレット | ツアー |
| 場所系 | 場所 イベント |
| 予算 | |
| 日程 | 出発月 出発日 帰国月 帰国日 期間 |
| ホテル | |
| 食事 | |
| 買物 | |
| 気候 | |
| 道具 | 交通手段 その他 |
| 相手 | パートナー 人数 |
| 関連手続き | パスポート チケット その他 |

図 3: ツアーに関する情報テーブルに含まれる情報

6. 指示対象の探索

指示対象の推定探索は、以下の順序で行われる。

- (1) 動詞が取りうる格の意味カテゴリ、修飾部、指示詞・形式名詞部の名詞句以外の名詞句、から指示対象探索のための制約を決定する。
制約には、次のようなものがある。

意味カテゴリ制約, 属性制約, 時間制約,
程度制約, 特徴制約, 否定制約

また、述語接続助詞で接続された文の場合
は、前の文からも制約が取り出される。

- (2) 決定された制約を満足する語を、文脈テーブルの焦点、他の語群、情報テーブルという順で探索していくことによって、指示対象を決定する。

探索の流れを以下の対話例で概略的に解説する。

[例]

U_n: こちらだと、バルセロナにも行きます。

U_{n+1}: さっきのたくさん行くやつでは、そこも行くのですか。

[1] U_n までの状態

U_n までの対話で、以下の 2 つのツアーに関する
ツアー・テーブルができているとする。

ツアー 1 → 訪問地が多い

ツアー 2 → バルセロナに行く

U_n における指示詞「こちら」は、ツアー 2 を指示
しているとする。この U_n までで保持される文脈テ
ーブルとツアー・テーブルの状態は図 4 の通りであ
る。

| | U ₀U _{n-1} | U _n |
|------|--------------------------------------|--------------------|
| (1) | | 肯定 |
| (2) | | 断定 |
| (3) | | 行く |
| (4) | | バルセロナ |
| (5) | | バルセロナ |
| (6) | | 場所 |
| (7) | | [ツアー2, ツアー名] |
| (8) | | [ツアー2, バルセロナ, 場所] |
| (9) | | [ツアー2, ツアー2, ツアー名] |
| (10) | | 担当者 |
| (11) | | n |

| | |
|-----------------------------------|---------------------------------|
| ツアー 1 場所: A, B, C, D . . | ツアー 2 場所: C, バルセロナ . . |
|-----------------------------------|---------------------------------|

図 4: U_n までの文脈テーブルと
ツアー・テーブルの状態

[2] U_{n+1} の解析

U_{n+1} における指示詞「そこ」と形式名詞「やつ」
の指示対象は、図 5 に示される方法で探索される。

| |
|---|
| その照応: 述部の動詞: 行く 意味カテゴリ制約: 場所, イベント, ↓ U _n の焦点「バルセロナ」より そこ=バルセロナ |
| やつの照応: 述部の動詞: 行く 意味カテゴリ制約: ツアー, パンフレット, 時間制約: さっき 程度制約: たくさん行く ↓ 情報テーブルの訪問地の多いツアーより やつ=ツアー1 |

図 5: 指示対象の照応

その結果、 U_{n+1} によって更新される文脈テーブルは、図6のようになる。更新された文脈テーブルの U_{n+1} においては、焦点はバルセロナのままで、「ツアー1で行くバルセロナは、ツアー2では行きますか」と解析されて情報が格納されることになる。

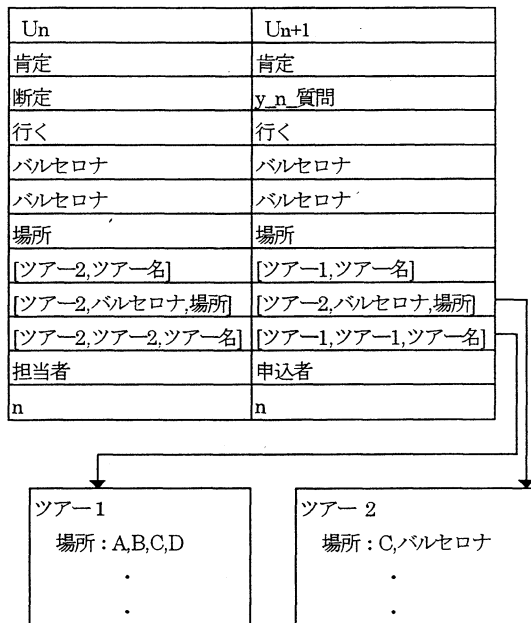


図6: 更新された文脈テーブル

7. 解析結果

開発したシステムを用いて、7対話例計176文に対して実験を行った。対話は、実際のパック旅行選択対話[1]を基に、簡潔化して変更を加えたものである。

まず、解析された対話全体において出現した指示詞・形式名詞の数は、それぞれ表1のようであった。

| | |
|------|-----|
| 指示詞 | 70 |
| 形式名詞 | 32 |
| 合計 | 102 |

表1: 対話中の指示詞・形式名詞の数

また、それらの指示詞・形式名詞における照応解析の正解率は、表2のようであった。

| | |
|------|-----|
| 指示詞 | 87% |
| 形式名詞 | 71% |
| 合計 | 82% |

表2: 指示詞・形式名詞の解析に対する正解率

表2より、全体として82%の指示詞・形式名詞の指示対象を解析できていることから、本論で提案した方法が有効であったことが示された。

8. 考察

システムが解析を誤ったものに対して調査した結果、今後の主な課題としては、以下のようなものがあげられる。

(1) 表層的な言語制約の強化

システムが解析を誤った理由の主なものに、抽出する制約が不足していることに起因しておいたものがある。そのような場合は、語用論的な制約等の表層的な言語制約を強化していけば、かなり精度をあげることができるのではないかと考えられる。本論で提案する方法では、表層的言語情報による制約が少なくとも、精度の高い指示詞・形式名詞の照応を実現することができたが、表層的な言語情報をもっと組み込めば、より強力なシステムになるだろうと思われる。

(2) 解析誤りに対して

システムは、それまでの文脈情報を用いて指示詞・形式名詞を解析し、その結果を用いて、更に文脈情報を更新し、保持していく。ということは、解析を誤った場合でも、誤ったまま、文脈として保持していつてしまう。

人間は、そのように、「勘違い」をしてしまった場合でも、対話が進行していくにつれて、「勘違い」を自分で修正していくことができる。しかし、これをコンピュータで解析するには、学習能力やより複雑な推論が必要になると考えられる。この点については、今後の課題として研究を進めなければならない点だと考えられる。

参考文献

- [1] 江原, 井ノ上, 他: ATR 音声対話データベースの内容, 1990, ATR 自動翻訳電話研究所
- [2] Marilyn A. Walker, Masayo Iida, and Sharon Cote: Centering in Japanese Discourse, 1990, In Proceedings of the 13th COLING
- [3] Marilyn A. Walker, Masayo Iida, and Sharon Cote: Japanese Discourse and the Process of Centering, 1994, Computational Linguistics 20-2